ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND MANAGEMENT



Check for updates

UDC 004.11.5
https://doi.org/10.23947/2687-1653-2022-22-1-67-75

Original article



Comparison of machine learning models for coronavirus prediction

Brou Kouame Amos ✉, **Ivan Smirnov** , **Mabouh Moise Hermann**
Peoples' Friendship University of Russia (RUDN) (Moscow, Russian Federation)
✉ broukouameamos9@gmail.com

Introduction. Coronavirus, also known as COVID-19, was first detected in Wuhan, China, in December 2019. It is a family of viruses ranging from the common cold to severe acute respiratory syndrome (SARS). The symptoms of such a virus are similar to those of a cold or seasonal allergies. Like other respiratory viruses, it is mainly transmitted through airborne droplets when coughing or sneezing. Therefore, the recognition of COVID-19 requires careful laboratory analysis, and the reduction of recognition resources is a major challenge. On 11 March, 2020, the World Health Organization (WHO) declared COVID-19, caused by SARS-CoV-2, a pandemic, as there had been an exponential increase in cases worldwide, and demand for intensive beds and related structures had far exceeded existing capacity. The first examples of this are the regions of Italy. Brazil registered the first case of SARS-CoV-2 on 02/26/2020. Transmission of the virus in this country shifted very quickly from imported cases to local and, finally, community missions, with the Brazilian federal government announcing national community transmission on 03/20/2020. As of March 23, in the state of São Paulo with a population of about 12 million people, where the Israelita Albert Einstein Hospital is located, 477 cases of the disease and 30 related deaths were registered, and on March 27, there were already 1223 cases of COVID-19 with 68 concomitant deaths. To slow the spread of the virus in the state of São Paulo, quarantines and social distancing measures were introduced. One of the motivations for this challenge is the fact that, in the context of an extensive healthcare system with the possible limitation of SARS-CoV-2 testing, it is not practical to test every case, and test results can only be used in testing the target subpopulation. The study objective is to build a model based on machine learning that can predict the detection of SARS-CoV-2 from medical data. For this, various classification models of machine learning are compared, and the best one to predict coronaviruses is determined. The comparison is based on individuals in class 1, i.e., those with a positive test. Therefore, it is required to determine the machine learning model with the best response and F1 score for class 1.

Materials and Methods. An open-source data set from the Israelita Albert Einstein Hospital in São Paulo, Brazil, was taken as a basis. The following machine learning models were used for the study: RandomForests (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and AdaBoost (AB), as well as the 10-time cross-validation technique. Some machine learning performance measures, such as accuracy, recall, and F1 score were evaluated.

Results. Out of a total of 5,644 people tested during the COVID-19 pandemic, 5,086 people tested negative and 558 people tested positive. At the same time, support for machine vectors showed the best results in detecting coronavirus with a recall of 75 % and an F1 score of 60 % compared to models: Random drill, KNN, LR, AB, and DT.

Discussion and Conclusions. It was found that when using AB algorithms, greater accuracy is achieved, but the stability of the LSVM algorithm is higher. Therefore, it can be recommended as a useful tool for detecting COVID-19.

***Keywords*:** COVID-19 detection, classification, machine learning models.

For citation: Brou Kouame Amos, I. Smirnov, Mabouh Moise Hermann. Comparison of machine learning models for coronavirus prediction. Advanced Engineering Research, 2022, vol. 22, no. 1, pp. 67–75. https://doi.org/10.23947/2687-1653-2022-22-1-67-75



Сравнение моделей машинного обучения для прогнозирования коронавируса

К. А. Бру ✉, **И. В. Смирнов**, **М. М. Эрманн**
Российский университет дружбы народов (Москва, Российская Федерация)
broukouameamos9@gmail.com

Введение. Коронавирус, также известный как COVID-19, впервые обнаружен в Ухане (Китай) в декабре 2019 г. Он представляет собой семейство вирусов, начиная от простуды и заканчивая тяжелым острым респираторным синдромом (ТОРС). Симптомы такого вируса схожи с симптомами простуды или сезонных заболеваний. Как и другие респираторные вирусы, он в основном передается воздушно-капельным путем во время кашля или чихания. Поэтому распознавание COVID-19 требует тщательного лабораторного анализа, а сокращение ресурсов распознавания является серьезной научной задачей. Всемирная организация здравоохранения (ВОЗ) 11.03.2020 объявила COVID-19, вызванный SARS-CoV-2, пандемией, поскольку во всем мире произошел экспоненциальный рост числа случаев заболеваний, а спрос на интенсивные койки и соответствующие структуры намного превысил существующие возможности. Первыми примерами этому являются регионы Италии. Бразилия зарегистрировала первый случай SARS-CoV-2 26.02.2020. Передача вируса в этой стране очень быстро перешла от завезенных случаев к местным и, наконец, общинным миссиям, а федеральное правительство Бразилии объявило о национальной общинной передаче 20.03.2020. В штате Сан-Паулу с населением около 12 млн человек, где находится больница Альберта Эйнштейна, по состоянию на 23.03.2020 зарегистрировано 477 случаев заболевания и 30 связанных с ними смертей, а 27.03.2020 имели место уже 1223 случая COVID-19 с 68 сопутствующими смертями. Для замедления распространения вируса в штате Сан-Паулу были введены карантин и меры социального дистанцирования. Одним из мотивов этой проблемы является тот факт, что в контексте обширной системы здравоохранения с возможным ограничением тестирования SARS-CoV-2 нецелесообразно тестировать каждый случай, а результаты тестов могут быть использованы при проверке только целевой субпопуляции. Целью работы является построение на основе машинного обучения модели, способной прогнозировать обнаружение SARS-CoV-2 по медицинским данным. Для этого проводится сравнение различных классификационных моделей машинного обучения и определяется лучшая из них с целью прогнозирования коронавирусов. Сравнение основано на лицах в классе 1, т. е. с положительным тестом. Поэтому необходимо определить модель машинного обучения с лучшим отзывом и F1-баллом для класса 1.

Материалы и методы. За основу принят набор данных с открытым исходным кодом из израильской больницы Альберта Эйнштейна в Сан-Паулу. Для исследования использованы модели машинного обучения: RandomForests (RF), K-ближайший сосед (KNN), Машина опорных векторов (SVM), Логистическая регрессия (LR), Дерево решений (DT) и AdaBoost (AB), а также 10-временная техника перекрестной проверки. Проведена оценка некоторых показателей производительности машинного обучения, таких как точность, отзыв и оценка F1.

Результаты исследования. Из 5644 человек, протестированных во время пандемии COVID-19, 5086 человек дали отрицательный результат и 558 человек — положительный. При этом поддержка машинных векторов показала лучшие результаты в обнаружении коронавируса с отзывом — 75 % и оценкой F1 — 60 % по сравнению с моделями: Random drill, KNN, LR, AB и DT.

Обсуждение и заключение. Установлено, что при использовании алгоритмов AB достигается большая точность, однако стабильность алгоритма LSVM является более высокой. Поэтому его можно рекомендовать как полезный нструмент для выявления COVID-19.

Ключевые слова: выявление COVID-19, классификация, модели машинного обучения.

***Для цитирования*:** К. А. Бру. Сравнение моделей машинного обучения для прогнозирования коронавируса / К. А. Бру, И. В. Смирнов, М. М. Эрманн // Advanced Engineering Research. — 2022. — Т. 22, № 1. — С. 67–75. https://doi.org/10.23947/2687-1653-2022-22-1-67-75

1. Introduction

The coronavirus is a very severe acute respiratory syndrome caused by the SARS-COV-2 virus. This virus, which can infect humans or animals, was discovered in the Chinese region of Wuhan, more precisely in the province of Hubei, during the pneumonia epidemic of January 2020 [1, 2]. It is therefore the seventh human coronavirus. To everyone's surprise, this virus spread worldwide, causing 318,599 deaths and 4,806,299 infected persons [3].

SARS-CoV-2, SARS-CoV and MERS-COV (Middle East Respiratory Syndrome Coronavirus) cause severe pneumonia with a mortality rate of 2.9 %, 9.6 % and 36 % respectively [4–6].

The other four viruses, namely OC43, NL63, HKU1, and 229E, are responsible for illnesses related to mild symptoms [7].

It should be noted that since the Covid-19 epidemic, there has been much speculation about the origin of this virus [8]. Some said that it was the result of work done in a laboratory. However, after studies conducted on genetic data, this hypothesis was dismissed [9]. Analysis and comparison with the genomes of previously known coronaviruses clearly show that SARS-COV-2 is different from other coronaviruses [8, 11]. The virus responsible for the coronavirus (SARS-COV-2) is similar to the SARS virus of bats [2]. Thus, the Covid-19 virus is believed to have originated from a bat coronavirus that became infectious to humans while acquiring genes specific to pangolin coronaviruses. It should be noted that the actual causes of Covid-19 are still unclear.

The symptoms of Covid-19 are similar to those of seasonal flu. The disease is more severe in the elderly and in people who are vulnerable to certain chronic diseases. Patients with Covid-19 can have symptoms ranging from mild to severe. The most common symptoms are fever (83 %), cough (82 %) and breathlessness (31 %) [12]. In patients with pneumonia, the X-ray of the lungs shows numerous mottles and ground glass opacity [12, 13].

Gastrointestinal symptoms associated with patients with Covid-19 include vomiting, diarrhoea, and abdominal pain [12, 14].

We also see a decrease in lymphocytes and eosinophils, lower haemoglobin levels, and an increase in white blood cells and neutrophils [15–18].

The manifestation of Covid-19 in children is different from that in adults. In children, the symptoms are mild. However, in some children, we have seen severe and fatal cases [19–27].

Like all other viruses, Covid-19 is transmitted mainly by the respiratory route. Among these routes of transmission, we have droplet transmission, which is the most widespread [28, 29]. Other transmission routes exist, namely the faecal route, via saliva. Indeed, SARS-CoV-2 RNA was found in the stool of a patient with Covid-19 [31]. SARS-CoV-2 RNA can be detected on inanimate surfaces (door handles). People who have been in contact with these surfaces could be contaminated [29].

This model will make it possible to identify positive and negative cases from the dataset studied and the elements responsible for COVID-19. The proposed prediction model ensures that it tracks the results regarding this epidemic situation so that the huge economic losses, the spread of the community, the amount of detachment social gens can be detected and a precise decision can also be made accordingly. This method will allow government authorities to put in place preventive measures based on our future work to predict the onset of this disease in the future.

2. Data Resources and Methods

The dataset used was uploaded to Kaggle. It is open source and available on this link kaggle.com/einsteindata4u/covid19. This dataset contains anonymized data in accordance with best international practices and patient recommendations at the Israelita Albert Einstein Hospital in São Paulo, Brazil. This section describes the proposed approach and a detailed overview of the tasks. These tasks can help to understand and extract knowledge from COVID 19 data, which can help countries contain the spread of the virus, raise awareness, launch initiatives, determine if mitigation has a positive effect or not, identify other factors affecting the virus, etc. This will allow countries to prepare for what may happen in the near future. This could help save lives and alleviate the agony. Epidemiological information includes various characteristics of the case studied, including case identification, age, sex, target value, lymphocytes, leukocytes, monocytes, hco3, etc.

2.1. Data Pre-processing

In data analysis, the most important step is pre-processing. However, it is not clear what methods of pre-treatment the author used. This part must be completed.

2.2. Data Transformation

The data is transformed to be processed and stored in. xls for further processing. All data were normalized to have a mean of zero and a unit standard deviation. With a dataset containing 111 characteristics, data mining eliminated missing values (78 characteristics) and retained important characteristics (33). This exploratory analysis of the data also allowed us to identify two categories of characteristics, namely virus-related characteristics and blood-related characteristics. The target value is divided into two categories which are negative cases coded by 0 and positive cases coded by 1.

The dataset from the Israelita Albert Einstein Hospital in São Paulo is divided into training and test data. 70 % of the data is used for predictive model training, and the remaining 30 % is used for testing. The objective of model training is to adapt the model using data from the training set. After the model is formed, the prediction models sound tested to evaluate performance in the test datasets.

2.3. The Proposed Models

This section describes the different machine learning models used in this paper. These models are: Random Drills (RF), K-plus Close Neighbors (KNN), Linear Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), and AdaBoost (AB).

Random Forest (RF)

Random forests (RF) or random decision forests were first proposed in 1995. This is a general classification training method that tends to work better than traditional decision tree classification methods (Gangaie et al., 2019). Decision trees are the fundamental RF classifiers that vote for each of the forecasts, and the survival prediction is based on the majority voting method in each tree (Breiman, 2001). The accuracy of each tree and the independence of the trees from each other provide the reliability of the classification. We used 100 trees to predict two target classes, survival or death of patients with hepatitis.

Nearest Neighbor (KNN)

The K-Nest Neighbor (KNN) classifier is one of the most commonly used classification algorithms. This algorithm can be used in several applications. It saves all valid attributes and classifies new attributes according to their similarity dimension. KNN is a statistical recognition model method for detecting the different classes of a model. A tree data structure is used to determine the distance between the point of interest and the points in the training dataset. The attribute is classified by its neighbors. In the classification method, the value of k is always a positive integer closest to the neighbor. The nearest visions are selected from a set of classes or property values of the object.

Support Vector Machine (SVM)

SVM-controlled learning method is used for classification and regression [29]. This algorithm is a relatively new approach and has performed well in recent years. The SVM classifier is based on linear classifiers and in the data separated by a row, the SVM isolates the objects in the specified classes. It can also identify and classify instances that are not supported by the data. The only extension of this algorithm is to perform a regression analysis to obtain a linear function, and another extension teaches to classify the elements to obtain a classification of individual elements.

Logistic Regression Model (LR)

Logistic regression is the corresponding regression analysis that should be performed when the dependent variable is dichotomous (binary). Like all regression analyses, logistic regression is predictive analysis. It is used to describe the data and explain the relationship between a dependent binary variable and one or more nominal, ordinal, interval or ordinal independent variables, report [30, 31]. This approach assumes that the binary result follows a binomial distribution.

Decision Tree (DT) Model

The Decision Tree is a controlled learning method that is used to solve classification and regression problems, but it is more used to solve classification. This is a powerful classification method for disease prediction. This is a tree model where the internal nodes represent the characteristics of a data set, the branches represent the decision rules, and each leaf node represents a result. The decision tree consists of two nodes, a decision node and a leaf node. Decision nodes have multiple branches and are used to make a decision, while leaf nodes are the result of those decisions.

Model AdaBoost (AB)

AdaBoost, short for "Adaptive Boosting", is the first boost algorithm proposed by Freund and Schapire in 1996. Its goal is to turn weak predictors into strong predictors to solve classification problems. For classification, the final equation can be put under the heading below:

$$F(x) = \text{sign}\left(\sum_{m=1}^{M} \theta_m \, f_{m(x)} \right) \quad (1)$$

Where f_m denotes the weak classifier m and m denotes the corresponding weight. AdaBoost can be used for face recognition, as it is a standard algorithm for detecting faces in images. AdaBoost is fast, requires no setup, and is simple and easy to program. Plus, it has the flexibility to be able to be combined with any machine learning algorithm.

2.4. Evaluation of Performance Measures

For the comparison of the different classification algorithms used in this paper, some metrics were evaluated. These are accuracy, recall, and F1-score. These metrics are calculated based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The standardized confusion matrix illustrates the relationship between classification results and predicted classes. The level of the classification performance is calculated by the number of samples correctly and incorrectly classified in each class.

The accuracy is calculated based on the total number of correct predictions, defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (2)$$

Recall, or sensitivity, is the proportion of true positive predictions that have been correctly identified, defined as follows:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

The F1 score is the harmonic mean of accuracy and recall, and it is calculated by:

$$\text{Score F1} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{4}$$

3. Result

The objective of this paper is to compare the different models of machine learning for the detection of coronavirus. Our task was to find out which machine learning model has the best recall and f1-score for Class 1. The learning machine models used are: Radom drill, k-nearest neighbor, logistic regression, support vector machine, AdaBoost, and decision tree. Out of a total of 5,644 people tested for COVID-19, 5,086 people tested negative and 558 people tested positive. The results of our study are presented in Figure 1 and Figure 3. These results show that the vector-machine gave better results with a recall of 75 % and an F1 score of 60 %. The different learning curves were also traced in order to understand the phenomenon of over-fitting and under-fitting Figure 2. Indeed, the learning curve is very well known to data scientists, the learning curve shows the efficiency and quality of learning of our machine learning model. Learning curves are widely used as a diagnostic tool in machine learning for algorithms that incrementally learn a training data set. This means that we increase our dataset by a certain step, and then we see the performance of our model. The model can be evaluated on the training dataset and on the exception validation dataset after each update during training, and it traces the measured performance. This can be represented as a curve.

```
RandomForest
[[91  4]
 [11  5]]
              precision    recall  f1-score   support

           0       0.89      0.96      0.92        95
           1       0.56      0.31      0.40        16

    accuracy                           0.86       111
   macro avg       0.72      0.64      0.66       111
weighted avg       0.84      0.86      0.85       111
```

```
AdaBoost
[[91  4]
 [ 9  7]]
              precision    recall  f1-score   support

           0       0.91      0.96      0.93        95
           1       0.64      0.44      0.52        16

    accuracy                           0.88       111
   macro avg       0.77      0.70      0.73       111
weighted avg       0.87      0.88      0.87       111
```

```
KNN
[[88  7]
 [ 8  8]]
              precision    recall  f1-score   support

           0       0.92      0.93      0.92        95
           1       0.53      0.50      0.52        16

    accuracy                           0.86       111
   macro avg       0.72      0.71      0.72       111
weighted avg       0.86      0.86      0.86       111
```

```
DecisionTree
[[86  9]
 [11  5]]
              precision    recall  f1-score   support

           0       0.89      0.91      0.90        95
           1       0.36      0.31      0.33        16

    accuracy                           0.82       111
   macro avg       0.62      0.61      0.61       111
weighted avg       0.81      0.82      0.81       111
```

```
Logistic_Regression
[[92  3]
 [10  6]]
              precision    recall  f1-score   support

           0       0.90      0.97      0.93        95
           1       0.67      0.38      0.48        16

    accuracy                           0.88       111
   macro avg       0.78      0.67      0.71       111
weighted avg       0.87      0.88      0.87       111
```

```
SVM
[[83 12]
 [ 4 12]]
              precision    recall  f1-score   support

           0       0.95      0.87      0.91        95
           1       0.50      0.75      0.60        16

    accuracy                           0.86       111
   macro avg       0.73      0.81      0.76       111
weighted avg       0.89      0.86      0.87       111
```

Fig. 1. Classification report of different machine learning models

Fig. 2. Learning curve of different machine learning models

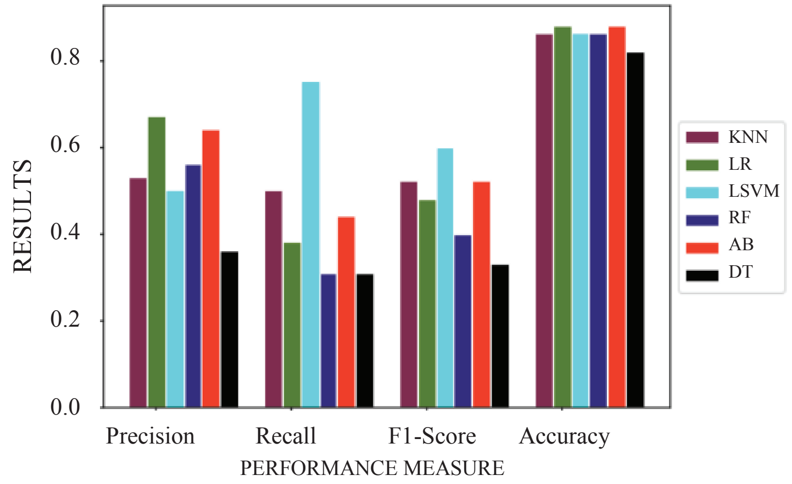


Fig. 3. Results of predictions from various machine learning techniques

Figure 3 shows the performance of the different machine learning algorithms according to the performance measures used in this paper. We see that for recall and F1-score, LSVM outperforms the other machine learning models used, namely LR, KNN, RF, AB, and DT. For accuracy, LR is much better than the others. As for accuracy, we find that LR and AB performed better than the other models. In this paper, we chose recall and F1 score to measure the performance of the model. Recall allowed us to correctly identify the Covid-19 positive test subjects among all the real positive cases. As for the F1 score, we used it because we had an imbalance between different classes, i.e., positive and negative cases.

4. Discussion and Conclusion

The data used in this paper was collected at the Israelita Albert Einstein Hospital in São Paulo, Brazil. After an exploratory analysis, two categories of characteristics were identified. These are the characteristics related to the virus and the characteristics related to the blood. Out of a total of 5,644 people tested with COVID-19, 5,086 people tested negative and 558 people tested positive. The results of this study clearly illustrated that in relation to our goal, machine vector support showed better results in coronavirus detection with a recall of 75 % and an F1 score of 60 %. This co-calculation was done with the other machine learning models, namely the Radom drill, the k-nearest neighbor, the logistic regression, the AdaBoost, and the decision tree. As such, this model can be useful for the diagnosis of COVID-19. However, it is possible to optimize the parameters of this model in order to improve its performance.

After the analysis of the learning curve in Figure 2, we find that apart from the supporting sensor, other machine learning models can be studied for the detection of COVID-19. These include AdaBoost and k-nearest neighbor. Indeed, we find that if we perform a little more advanced optimization of the parameters of these models, they could be candidates for the diagnosis of COVID-19 because the difference between the learning score curve and the validation score curve would have reduced the model's ability to generalize.

References

1. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–273. https://doi.org/10.1038/s41586-020-2012-7

2. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 265–269. https://doi.org/10.1038/s41586-020-2008-3

3. World Health Organization Coronavirus Disease 2019 (COVID-19) Situation Report-97. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200426-sitrep-97-covid-19.pdf

4. Wang C, Horby PW, Hayden FG, et al. A novel coronavirus outbreak of global health concern. Lancet. 2020;395:470–473. https://doi.org/10.1016/S0140-6736(20)30185-9

5. Hui DSC, Zumla A. Severe acute respiratory syndrome: historical, epidemiologic, and clinical features. Infect Dis Clin North Am. 2019;33:869–889. https://doi.org/10.1016/j.idc.2019.07.001

6. Azhar EI, Hui DSC, Memish ZA, et al. The Middle East respiratory syndrome (MERS). Infect Dis Clin North Am. 2019;33:891–905. https://doi.org/10.1016/j.idc.2019.08.001

7. Corman VM, Muth D, Niemeyer D, et al. Hosts and sources of endemic human coronaviruses. Adv Virus Res. 2018;100:163–188. https://doi.org/10.1016/bs.aivir.2018.01.001

8. Andersen KG, Rambaut A, Lipkin WI, et al. The proximal origin of SARS-CoV-2. Nat Med. 2020;26:450–452. https://doi.org/10.1038/s41591-020-0820-9

9. Almazán F, Sola I, Zuñiga S, et al. Coronavirus reverse genetic systems: infectious clones and replicons. Virus Res. 2014;189:262–270. https://doi.org/10.1016/j.virusres.2014.05.026

10. Nao N, Yamagishi J, Miyamoto H, et al. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. mBio. 2017;8:e02298. http://dx.doi.org/10.1128/mBio.02298-16

11. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

12. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA. 2020;323:1061. https://doi.org/10.1001/jama.2020.1585

13. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727–733. https://doi.org/10.1056/NEJMoa2001017

14. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395:507–513. https://doi.org/10.1016/S0140-6736(20)30211-7

15. Lippi G, Plebani M. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks. Clin Chem Lab Med. 2020;58:1063–1069. https://doi.org/10.1515/cclm-2020-024

16. Bhargava A, Fukushima EA, Levine M, et al. Predictors for severe COVID-19 infection. Clin Infect Dis. 2020;71:1962-1968 https://doi.org/10.1093/cid/ciaa674

17. Wang CZ, Hu SL, Wang L, et al. Early risk factors of the exacerbation of coronavirus disease 2019 pneumonia. J Med Virol. 2020;91:2593-2599 https://doi.org/10.1002/jmv.26071

18. Hamming I, Timens W, Bulthuis ML, et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol. 2004;203:631–637. https://doi.org/10.1002/path.1570

19. Renu K, Prasanna PL, Valsala Gopalakrishnan A. Coronaviruses pathogenesis, comorbidities and multi-organ damage — a review. Life Sci. 2020;255:117839. https://doi.org/10.1016/j.lfs.2020.117839

20. Long B, Brady WJ, Koyfman A, et al. Cardiovascular complications in COVID-19. Am J Emerg Med. 2020;38 :1504-1507 https://doi.org/10.1016/j.ajem.2020.04.048

21. Ruan Q, Yang K, Wang W, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med. 2020;46:846–848. https://doi.org/10.1007/s00134-020-05991-x

22. Lippi G, Favaloro EJ. D-dimer is associated with severity of coronavirus disease 2019: a pooled analysis. Thromb Haemost. 2020;120:876–878. http://dx.doi.org/10.1055/s-0040-1709650

23. Lang J, Yang N, Deng J, et al. Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. Plos One. 2011;6:e23710. https://doi.org/10.1371/journal.pone.0023710

24. Vicenzi E, Canducci F, Pinna D, et al. Coronaviridae and SARS-associated coronavirus strain HSR1. Emerging Infect Dis. 2004;10:413–418. https://doi.org/10.3201/eid1003.030683

25. Belen-Apak FB, Sarialioglu F. The old but new: can unfractioned heparin and low molecular weight heparins inhibit proteolytic activation and cellular internalization of SARSCoV2 by inhibition of host cell proteases? Med Hypotheses. 2020;142:109743. https://doi.org/10.1016/j.mehy.2020.109743

26. Henry BM, Benoit SW, Santos de Oliveira MH, et al. Laboratory abnormalities in children with mild and severe coronavirus disease 2019 (COVID-19): a pooled analysis and review. Clin Biochem. 2020;81:1–8. https://doi.org/10.1016/j.clinbiochem.2020.05.012

27. Sanna G, Serrau G, Bassareo PP, et al. Children's heart and COVID-19: Up-to-date evidence in the form of a systematic review. Eur J Pediatr. 2020;179:1079-1087 https://doi.org/10.1007/s00431-020-03699-0

28. Leung NHL, Chu DKW, Shiu EYC, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nature Med. 2020;26:676–680. https://doi.org/10.1038/s41591-020-0843-2

29. Abdi MJ, Giveki D. Automatic detection of erythemato-masquamous diseases using PSO-SVM based on association rules. Technical applications of artificial intelligence. 2013;26:603-608. https://doi.org/10.1016/j.engappai.2012.01.017

30. McDonald JH. Handbook of Biological Statistics, 3rd ed. Sparky House Publishing: Sparky House Publishing; 2014.

31. Mangiafico SS. An R companion for the handbook of biological statistics, 1.3.3 ed. New Brunswick, NJ: Rutgers Cooperative Extension; 2015.

Received 14.02.2022
Revised 28.02.2022
Accepted 01.03.2022

About the Authors:

Brou Kouame Amos, postgraduate of the Information Technology Department, Peoples' Friendship University of Russia (RUDN) (6, Miklikho-Maklaya St., Moscow, 117198, RF), ORCID, broukouameamos9@gmail.com

Smirnov, Ivan V., associate professor of the Information Technology Department, Peoples' Friendship University of Russia (RUDN) (6, Miklikho-Maklaya St., Moscow, 117198, RF), Cand.Sci. (Phys.-Math.), associate professor, Scopus, ORCID, smirnov-iv@rudn.ru

Mabouh Moise Hermann, postgraduate of the Information Technology Department, Peoples' Friendship University of Russia (RUDN) (6, Miklikho-Maklaya St., Moscow, 117198, RF), ORCID, mrmabouhmoise@gmail.com

Claimed contributorship

Brou Kouame Amos: basic concept formulation; research objectives and tasks; data pre-processing; analysis of research results. Mabouh Moise Hermann: text preparation; formulation of conclusions; data collection; the text revision. I. V. Smirnov: work control; the text revision; correction of the conclusions.

All authors have read and approved the final manuscript.

Об авторах:

Бру Куамэ Амос, аспирант кафедры «Информационные технологи» Российского университета дружбы народов (117198, РФ, г. Москва, Миклухо-маклая 6), ORCID, broukouameamos9@gmail.com

Иван Валентинович Смирнов, доцент кафедры «Информационные технологии», Российского университета дружбы народов» (117198, РФ, г. Москва, Миклухо-маклая 6), кандидат физико-математических наук, доцент, Scopus, ORCID, smirnov-iv@rudn.ru

Мабу Моисе Эрманн, аспирант кафедрой «Информационных Технологий», РУДН «Российский университет дружбы народов» (117198, РФ, г. Москва, Миклухо-маклая 6), ORCID, mrmabouhmoise@gmail.com

Заявленный вклад соавторов:

К. А. Бру — формирование основной концепции, цели и задачи исследования, предварительная обработка данных и анализ результатов исследований. М. М. Эрманн — подготовка текста, формирование выводов, сбор данных и доработка текста. И. В. Смирнов — контроль за работой, доработка текста и корректировка выводов.

Все авторы прочитали и одобрили окончательный вариант рукописи